

问题描述

Q: Hive的体系架构?

解决方法

A:

I 用户接口模块:

Ø 用来实现对Hive的访问。CLI是Hive自带的命令行界面; HWI是Hive的一个简单网页界面; JDBC、ODBC以及Thrift Server可向用户提供进行编程的接口, 其中Thrift Server是基于Thrift软件框架开发的, 提供Hive的RPC通信接口。

I 驱动模块 (Driver) :

Ø 负责把HiveQL语句转换成一系列MR作业, 所有命令和查询都会进入驱动模块, 通过该模块的解析变异, 对计算过程进行优化, 然后按照指定的步骤执行。

Ø 解析器: 将SQL字符串转换成抽象语法树AST, 这一步一般都是用第三方工具库完成, 比如antlr; 对AST语法树进行分析, 比如表是否存在、字段是否存在、SQL语义是否有误。

Ø 编译器: 将AST编译生成逻辑执行计划。

Ø 优化器: 对逻辑执行计划进行优化。

Ø 执行器: 把逻辑执行计划转换成物理执行计划。对于hive来说, 就是MR/TEZ/Spark;

I 元数据存储模块 (Metastore) :

Ø 通常与MySQL数据库连接后创建的一个MySQL实例, 也可以是Hive自带的Derby数据库实例。此模块主要保存表模式和其他系统元数据, 如表的名称、表的列及其属性、表的分区及其属性、表的属性、表中数据所在位置信息等。

I Thrif server :

Ø 是Facebook开发的一个软件框架, 它用来进行可扩展且跨语言的服务开发, hive集成该服务, 能让不同的编程语言调度HIVE接口。